



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **11212728 A**

(43) Date of publication of application: 06 . 08 . 99

(51) Int. Cl

G06F 3/06
G06F 3/06

(21) Application number: **10012457**

(22) Date of filing: 26 . 01 . 98

(71) Applicant: **HITACHI LTD**

(72) Inventor: **ISHIKAWA ATSUSHI**
MATSUMOTO YOSHIKO
TAKAMOTO KENICHI

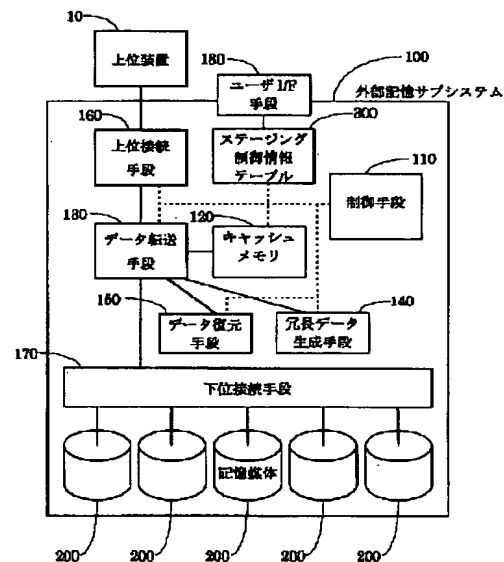
(54) EXTERNAL STORAGE SUB-SYSTEM

(57) Abstract:

PROBLEM TO BE SOLVED: To provide an external storage sub-system for guaranteeing maximum response time and accelerating single/multiple sequential read.

SOLUTION: A staging control information table 300 for specifying a range for performing staging from a storage medium 200 to a cache memory 120 is provided and a data transfer means 130 performs a staging processing to the cache memory 120 corresponding to the instruction of the staging control information table 300. By simultaneously reading redundant data as well at the time of reading data from the storage medium 200, response time in the case of failing in data read is shortened. Also, since a look-ahead processing is performed without performing look-ahead judgement, the sequential read is accelerated. Further, when a look-ahead direction is specified by the staging control information table 300, since look-ahead in a reverse direction is executed, a high-speed processing is made possible even at the time of the reverse reproduction of the data.

COPYRIGHT: (C)1999,JPO



BEST AVAILABLE COPY

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-212728

(43) 公開日 平成11年(1999) 8月6日

(51) Int.Cl.⁹

G 0 6 F 3/06

識別記号

3 0 2

5 4 0

F I

G 0 6 F 3/06

3 0 2 A

5 4 0

審査請求 未請求 請求項の数 9 O L (全 12 頁)

(21) 出願番号 特願平10-12457

(22) 出願日 平成10年(1998) 1月26日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目 6 番地

(72) 発明者 石川 篤

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(72) 発明者 松本 佳子

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(72) 発明者 ▲高▼本 賢一

神奈川県小田原市国府津2880番地 株式会
社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 小川 勝男

(54) 【発明の名称】 外部記憶サブシステム

(57) 【要約】

【課題】 最大応答時間を保証、および、単一／多重シーケンシャル読み出しの高速化を実現する外部記憶サブシステムを提供する。

【解決手段】 記憶媒体からキャッシュメモリにステージングする範囲を指定するステージング制御情報テーブルをもち、ステージング制御情報テーブルの指示にしたがってデータ転送手段がキャッシュメモリにステージング処理をおこなう。記憶媒体からのデータ読み出し時に冗長データも同時に読み出すことにより、データ読み出しに失敗した場合の応答時間を短縮できる。また、先読み判定をおこなわずに先読み処理をおこなうため、シーケンシャルな読み出しを高速化できる。さらに、ステージング情報テーブルによって先読み方向を指定すれば、逆方向への先読みを実施できるので、データの逆再生時にも高速な処理が可能となる。

図 6

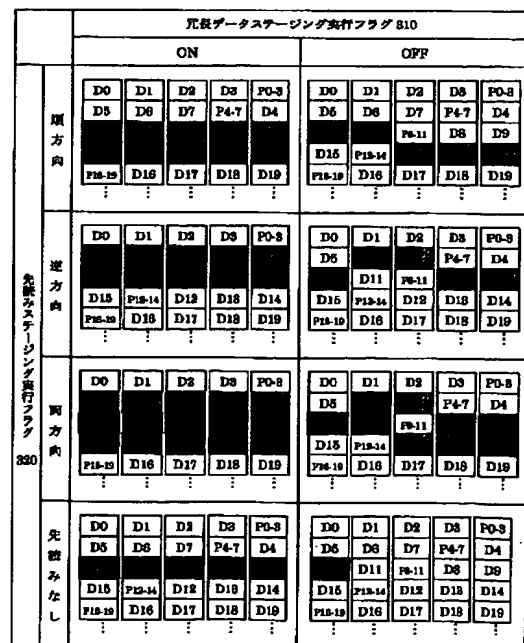


図 6 図 6 がキャッシュメモリ 120 へのステージング範囲 800。

【特許請求の範囲】

【請求項1】所定の単位に分割した複数のデータを記憶する記憶媒体と、上位装置に接続されて前記記憶媒体とのデータ転送を制御するデータ転送手段と、前記上位装置と前記記憶媒体との間で転送されるデータを一時的に記憶するキャッシュメモリと、複数のデータから冗長データを生成する冗長データ生成手段と、複数のデータと冗長データとから元のデータを復元するデータ復元手段と、前記冗長データと該冗長データを生成した複数のデータとをバリエーションとして管理する制御手段と、前記データを記憶媒体から読み出すときに、該データの冗長データあるいは該データに連続するデータを読み出すかどうかに関する情報を記憶するステージング情報テーブルとを有する外部記憶サブシステム。

【請求項2】前記ステージング情報テーブルは、記憶媒体から読み出すデータの冗長データを読み出すかどうかを決定する冗長データステージング実行フラグを含むことを特徴とする請求項1記載の外部記憶サブシステム。

【請求項3】前記冗長データステージング実行フラグが冗長データのステージングを指示している場合は、記憶媒体からデータを読み出すときに、該データが含まれるバリエーションのデータを読み出すことを特徴とする請求項2記載の外部記憶サブシステム。

【請求項4】前記ステージング情報テーブルは、記憶媒体から読み出すデータに連続するデータを読み出すかどうかを決定する先読みステージング実行フラグを含むことを特徴とする請求項1記載の外部記憶サブシステム。

【請求項5】前記先読みステージング実行フラグが順方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と同じ方向に連続したデータを読み出すことを特徴とする請求項4記載の外部記憶サブシステム。

【請求項6】前記先読みステージング実行フラグが逆方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と反対の方向に連続したデータを読み出すことを特徴とする請求項4記載の外部記憶サブシステム。

【請求項7】前記先読みステージング実行フラグが両方向の先読みを指示している場合、記憶媒体からデータを読み出すときに、データの並び順と同じ方向及び反対の方向に連続したデータを読み出すことを特徴とする請求項4記載の外部記憶サブシステム。

【請求項8】前記ステージング情報テーブルに冗長データステージング実行テーブルあるいは先読みステージングテーブルの内容を入力する手段を有することを特徴とする請求項1記載の外部記憶サブシステム。

【請求項9】前記上位装置から冗長データステージング実行フラグあるいは先読みステージング実行フラグに関する情報を受け取ったとき、前記制御手段は、前記ステージング情報テーブルに冗長データステージング実行フ

ラグあるいは先読みステージング実行フラグの内容を入力する手段を有することを特徴とする請求項1記載の外部記憶サブシステム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、コンピュータ装置に接続される外部記憶サブシステムに係わり、特に応答時間の保証が要求される外部記憶サブシステム、および、多重のシーケンシャルな読みだしにおいて高速なデータ転送能力を要求される外部記憶サブシステムに関する。

【0002】

【従来の技術】 1. 応答時間

従来は、RAIDなどの冗長データによるデータ復元機能を有する外部記憶サブシステムにおいて、記憶媒体からキャッシュメモリへのデータD10の読み出し（以下、ステージングと称す）に失敗した場合、データ復元に必要な他のデータと冗長データをステージングしてデータを復元した後に、上位装置にデータ転送することにより信頼性を向上させていた。

【0003】そのため、ステージングが正常終了した場合の応答時間、

ステージング正常終了時の応答時間＝当該データステージング時間＋データ転送時間＋その他処理時間

に比べ、ステージングに失敗した場合には、

ステージング失敗時の応答時間＝当該データステージング時間＋冗長データステージング時間＋データ転送時間＋データ復元時間＋その他処理時間

となり、応答時間が長くなっていた。

【0004】特開平7-200191号公報には、データ復元の必要性をいち早く認識して、上位装置からの処理を遅延することなしに予備記憶媒体へのデータを復元することを可能とするディスクアレイ装置に関して述べられている。

【0005】しかし、このディスクアレイ装置では、ステージングに失敗したI/O要求自身の高速化に関しては考慮されていなかった。

【0006】 2. シーケンシャルリード

外部記憶サブシステムでは、シーケンシャルリードを高速化するために、キャッシュメモリを設け先読み処理をおこなうことが一般的となっている。低速な記憶媒体からのデータを上位装置からの要求に先立ってキャッシュメモリに転送しておくこと（先読み処理）によって、上位装置からの読み出し要求を高速なキャッシュメモリからのデータ転送のみで済ませることができ、高速に処理することが可能となる。

【0007】従来の外部記憶サブシステムでは、一般に単一のシーケンシャルなリード要求に関しては、連続または同一の領域に対して数回の読み出しを受領したことにより上位装置からシーケンシャルなリード要求を受領

していると判断（シーケンシャル判定）した場合に、先読み処理をおこなって処理の高速化を図っていた。また、複数のシーケンシャルリード要求に関しては、シーケンシャル判定に要する情報を複数セット持ち、先読み処理を多重に動かすことにより処理を高速化していた。

【0008】しかし、これら技術においては、シーケンシャル判定に要する時間短縮や、シーケンシャル判定に要する情報のセット数により発生する先読み処理多重度の制限をなくすための考慮がされていなかった。

【0009】

【発明が解決しようとする課題】近年の情報化社会にともない、記憶装置の分野においてもマルチメディア対応に対する要求が高まっている。中でも、ビデオオンデマンドシステムやインターネットサーバシステムなどの画像／音声データを扱うシステムにおいては、サーバは記憶装置から大量の画像／音声などの情報を読みだし、単一または複数のクライアント（利用者）に提供する。

【0010】上記の画像／音声データを扱うシステムでは、データ読みだしの遅れが画像／音声の乱れや遅延の発生につながるため、記憶装置は最大応答時間を保証することが求められる。しかし、RAIDなどの冗長データによるデータ復元手段を有する外部記憶サブシステムにおいて記憶媒体からのデータの読み出しに失敗した場合、記憶媒体に再びアクセスして冗長データを読み出してから冗長データからデータを再生して上位装置に転送するため、応答時間が長くなってしまいう問題点があった。

【0011】また、上記のシステムで扱われる画像／音声のデータは一般に順次性のある（シーケンシャルな）データであることが多く、記憶装置はシーケンシャルな読みだしに対する高速処理も要求される。特に、ビデオオンデマンドシステムでは、画像の再生／逆再生などが行われるため、データの順方向のシーケンシャル読み出しにも、逆方向のシーケンシャル読み出しにも対応が必要である。さらに、現在のネットワーク環境におけるシステム運用を考えると、これらのシステムは多数のクライアントから同時にアクセスされることが多く、多数のシーケンシャルリード処理を平行して高速に処理することが不可欠となっている。

【0012】従来の外部記憶サブシステムでは、単一のシーケンシャルな読み出しに対しては、連続または同一の領域に複数の読み出しを受領した場合に後続領域に対する先読み処理を開始する事により処理を高速化し、複数多重のシーケンシャルな読み出しに対しては、シーケンシャル判定に用いる情報を複数セット持ち先読み処理を多重に動作させて高速化する方式が用いられることが多かった。しかし、この方式では先読み処理の多重度はシーケンシャル判定に用いる情報のセット数までに制限され制限以上のシーケンシャル読み出しが多重した場合には性能が低下してしまうことや、判定処理によるオー

バヘッドで性能が低下してしまうことが問題となっていた。

【0013】本発明の目的は、記憶媒体からのデータ読み出しに失敗した場合にも短い時間での応答を可能にし、さらに、単一または複数多重の順方向／逆方向のシーケンシャルな読み出しの処理能力を向上させ、シーケンシャルリードの多重度に対する制限を解消する外部記憶サブシステムを提供することである。

【0014】

10 【課題を解決するための手段】上記の目的を実現するために、本発明による外部記憶サブシステムでは、データ転送手段により記憶媒体からキャッシュメモリにステージングされる領域をユーザが指定する為に用いるステージング制御情報テーブルを設ける。

【0015】上位装置からリード要求を受領した場合、データ転送手段はステージング制御情報テーブルのユーザ指示に従って記憶媒体からキャッシュメモリにステージングする範囲を決定し、ステージング処理をおこなう。このステージング制御情報テーブルには、データ復元に必要な冗長データのステージングの有無、および、データD10に連続する領域に対する先読みステージングの有無、先読みの方向（順方向／逆方向／両方向）、先読みをおこなう量を設定する。このステージング制御情報テーブルは、上位装置や内蔵または外部接続されたユーザインタフェース手段を用いて設定／変更が可能であり、ユーザは外部記憶サブシステムを使用するシステムの読み出し要求の特性にあわせてキャッシュメモリへのステージング範囲を変更できる。

30 【0016】このステージング制御情報テーブルに、冗長データステージングの実行を指示することによって、上位装置からのリード要求時に、データ転送手段はリード要求のあった当該データとともにその冗長データもキャッシュメモリにステージングする。これにより、当該データの読み出しに失敗した場合でも、ステージング済の冗長データからデータを復元して速やかにデータを送信することができるため、当該データのステージングに失敗した場合の応答時間を改善できる。この制御方式を用いることにより応答時間を

40 《改善前》：データステージング時間＋冗長データステージング時間＋データ復元時間＋その他時間から

《改善後》：データステージング時間＋データ復元時間＋その他時間と冗長データステージング時間の分だけ短縮することができる。

50 【0017】また、ステージング制御情報テーブルに先読みステージング処理実行／先読み方向と先読み量を指示すれば、上位装置からの読み出し要求受領時に当該データに順方向／逆方向に連続する領域もキャッシュメモリにステージングする先読みステージング処理がおこな

われる。この方式では、先読み判定をおこなわないため、判定によるオーバーヘッドを回避して、シーケンシャルリード処理を高速化できる。また、先読み判定の情報も不要となり先読み判定情報の数による先読み多重度の制限がなくなるため、何多重でも先読みを動かすことが可能となる。また、逆方向への先読みを実施することにより、データの逆再生時にも高速な処理が可能となる。

【0018】

【発明の実施の形態】以下、本発明の実施形態の1例を図面を用いて説明する。

【0019】本実施例では5台の記憶媒体にRAID5方式で冗長化してデータを記録する例を示すが、記憶媒体の台数、データ冗長化／復元の方式、データと冗長データの配置などは、任意であり、本実施例以外の方式でもよい。

【0020】図1は、本発明による外部記憶サブシステムの構成の概略をあらわす図である。外部記憶サブシステム100は上位装置10に接続されており、データを格納する記憶媒体200、装置全体の制御を行う制御手段110、上位装置10とのデータ転送に用いられるキャッシュメモリ120、外部記憶サブシステム100内部のデータ転送や上位装置10とのデータ転送を制御するデータ転送手段130、冗長データを生成する冗長データ生成手段140、冗長データからデータを復元するデータ復元手段150、上位装置10とのインタフェースを制御する上位接続手段160、下位の記憶媒体とのインタフェースを制御する下位接続手段170、記憶媒体200からキャッシュメモリ120へのデータ転送

(以下、ステージングとよぶ)の制御に用いるステージング制御情報テーブル300及びユーザがステージング制御情報テーブル300などの装置設定を変更するためのインタフェース部として内蔵または接続されているユーザインタフェース手段180を有している。

【0021】この実施例に示した構成はあくまでも1例であり、別の構成による実現も可能である。

【0022】また、シーケンシャルリードの高速化のみを実現する場合には、冗長データ生成手段140やデータ復元手段150などのデータ冗長化に関する機能のない構成も可能である。

【0023】図2は、記憶媒体200におけるデータと冗長データの配置を示した図である。ここでは、5台の記憶媒体がRAID5方式により冗長化された場合のデータおよび冗長データの配置例を示している。

【0024】データはストライプと呼ばれる管理単位に分割される。冗長データ生成手段140では4つのデータストライプから1つの冗長データストライプを生成する。例えば、データストライプD0～D3から冗長データストライプP0～3を生成し、データストライプD4～D7から冗長データストライプP4～7を生成する。そしてデータストライプおよび冗長データストライプは

5台の記憶媒体200を巡回するように配置する。このような4つのデータストライプと1つの冗長データストライプの組はパリティグループと呼ばれ、データ復元手段150を用いれば、同一のパリティグループの他の3つのデータストライプと1つの冗長データストライプから、残り1つのデータストライプを復元する事が可能である。例えば、データストライプD2は、同一のパリティグループPG0～3の3つのデータストライプD0／D1／D3と冗長データストライプP0～3から復元できる。

【0025】先にも述べたように、記憶媒体の台数、データ冗長化／復元方式、データと冗長データの配置などは、この実施例以外の方法でも構わない。また、本発明によるシーケンシャルリードの高速化のみを実現する場合には、データ冗長化のないデータ配置も可能である。

【0026】図3では、ステージング制御情報テーブル300の内容について説明する。

【0027】ステージング制御情報テーブル300は、上位装置10からのリード要求に対しキャッシュメモリ120にステージングする範囲を指定する情報を格納するテーブルであり、冗長データステージング実行フラグ310、先読みステージング実行フラグ320、順方向先読み量330、逆方向先読み量340からなる。冗長データステージング実行フラグ310は、上位装置10から要求されたデータをステージングする際に当該データD10を復元するときに必要なデータ(本実施例では、同一パリティグループの他のデータと冗長データ)のステージング(以下、冗長データステージングとよぶ)を行うかどうか制御する為に用いられ、ON/OFFが設定され、ONの場合には冗長データステージングを実行する指示されているものとみなす。先読みステージング実行フラグ320は、上位装置10から要求されたデータをステージングする際に当該データD10に連続する領域もステージング(以下、先読みステージングとよぶ)の実行と先読みの方向を制御するために用いられ、順方向の先読み、逆方向の先読み、両方向の先読み、先読みなしの4つを指定できるものとする。順方向先読み量330には順方向に先読みステージングする先読み量、逆方向先読み量340には逆方向に先読みステージングする先読み量を設定でき、キャッシュメモリ120の容量を最大値に制限される。本実施例で示した制御情報テーブルは、あくまでも、発明を実現するためのテーブル構成の1例であり、各情報の持ち方・設定値などの違う別のテーブル構成も可能である。

【0028】ステージング情報管理テーブルは上位装置10またはユーザインタフェース手段180を通して指定できるので、ユーザは使用するシステムに合わせてステージング範囲を指定できる。上位装置10からステージング制御情報テーブル300を指定するときは、上位装置からステージング情報管理テーブルの内容に関する

パラメータを有するSCSIのMODE SELECT コマンドやベンダユニークのコマンドなどを発行し、コマンドの受領を認識した制御手段によりステージング情報管理テーブル300を更新する方法があげられる。ユーザインタフェース手段180を通してステージング制御情報300を指定する方法の例としては、ユーザインタフェース手段180から上位装置と同様なMODE SELECT コマンドやベンダユニークコマンドを発行する方法や、ユーザインタフェース手段180が直接ステージング制御情報テーブル300を書き換える方法、ユーザインタフェース手段180としてディップスイッチを用いてステージング制御情報を設定する方法などがある。これらのステージング制御情報テーブル300の設定方法も例であり、他の設定手段も可能である。

【0029】また、本実施例では、簡単のためにステージング制御情報テーブル300が外部記憶サブシステム100に1つしかない場合について示したが、外部記憶サブシステム100内がRAIDグループなどで領域分割されている場合には分割領域毎にステージング制御情報テーブル300を用意し、領域毎にステージング方式を変更することも可能である。また、外部記憶サブシステム100に複数の上位装置10が接続される場合には、上位装置10毎にステージング制御情報テーブル300を用意し、上位装置10毎にステージング方式を変更することも可能である。

【0030】さらに、本実施例では、冗長データステージングと先読みステージングの両方を実行可能な例を示したが、冗長ステージングだけあるいは先読みステージングだけが実行可能なテーブル構成も可能である。冗長データステージングのみを可能とする場合、ステージング制御情報テーブル300は冗長データステージング実行フラグ310のみで構成される。先読みステージングのみを可能とする場合、ステージング制御情報テーブル300は先読みステージング実行フラグ320および順方向先読み量330、逆方向先読み量340のみで構成される。

【0031】本実施例に上位装置10からのI/O要求が発行された場合の処理方式の1例を図4～図7を用いて説明する。ここでは、上位装置10からデータストライプD10にI/O要求が来た場合について説明するが、他のデータストライプにI/O要求が来た場合でも同様の処理となる。

【0032】図4は、本実施例におけるI/O処理の全体の概略を示すフローチャートである。上位装置10からデータストライプD10のリード要求を受領した(ステップ401)場合、制御手段110はまず上位装置10からのI/O要求種別の判定を行い(ステップ402)、リード要求であることを認識すると、次にキャッシュメモリ上に当該データD10がステージングされているかどうか確認する(ステップ403)。既にキャッ

シュメモリに当該データD10がステージングされている場合には、ステージング処理など(ステップ404～408)を行わずに、キャッシュメモリ120から上位装置10に当該データD10を転送し(ステップ409)、上位装置10にI/O処理の完了を報告して(ステップ410)処理を完了する。

【0033】また、キャッシュメモリに当該データD10がステージングされていない場合は、当該データD10をステージング実行中かどうか確認(ステップ404)し、当該データD10のステージング処理を実行中ならば実行中のステージング処理が完了するまで待ち(ステップ405)、当該データD10のステージング処理を未実行ならばステージング処理を実行する(ステップ406)。

【0034】ステージング処理完了後、当該データD10のステージングが正常に終了したかどうか判定する(ステップ407)。ここで、当該データD10のステージング失敗の場合にはデータ復元処理で当該データD10の復元を行う(ステップ408)。その後、キャッシュメモリ120から上位装置10に当該データD10を転送し(ステップ409)、上位装置10にI/O処理の完了を報告して(ステップ410)処理が完了となる。

【0035】上位装置10からライト要求を受領した(ステップ401)場合には、I/O要求種別判定でライト処理である事を認識(ステップ402)したあと、ライトに必要な処理(ステップ420)を行った後、上位装置10にI/O処理の完了を報告する(ステップ410)。

【0036】図5は、ステージング処理(ステップ406)を詳細に説明したフローチャートである。ステージング処理(ステップ406)では、まず、ステージング制御情報テーブル300の設定値を元にステージング範囲600を決定する(ステップ501)。冗長データステージング実行フラグ310と先読みステージング実行フラグ320の組合せとステージング範囲600の関係については、図6で1例を説明する。

【0037】次に、ステージング範囲600のデータストライプ分だけループして(ステップ502)、キャッシュ上に書くデータストライプがステージングされているかどうか確認し(ステップ503)、キャッシュ上にデータストライプがステージングされていない場合には、記憶媒体200にデータストライプの読み出し要求を発行する。

【0038】図6に、冗長データステージング実行フラグ310及び先読みステージング実行フラグ320の設定値とステージングされる領域の関係をまとめる。

【0039】本実施例では、冗長データステージング実行フラグ310の設定値はON/OFF、先読みステージング実行フラグ320の設定値は順方向、逆方向、両

方向、先読みなしの4種類が設定でき、順方向先読み量330および逆方向先読み量340にはそれぞれ4データストライプ分の先読みが指示されている場合を取り上げたが、テーブル構造や設定値が異なる場合でも同様の方法でステージング範囲600を決定する。また、上位装置10からデータストライプD10にリード要求が発行された場合であるが、他のデータストライプにリード要求が発行された場合に関しても同様である。また、本実施例では、順方向先読み量330に4データストライプ分の先読みを実行されるように指示されているものとして説明しているが、本実施例以外の方法による順方向先読み量330の設定も可能である。

【0040】冗長データステージング実行フラグ310がON、先読みステージング実行フラグ320が順方向と設定されている場合、ステージング範囲600は当該データD10を含むパリティグループPG8-11とそれに順方向に連続するパリティグループPG12-15となる。

【0041】冗長データステージング実行フラグ310がON、先読みステージング実行フラグ320が逆方向と設定されている場合、ステージング範囲600は当該データD10を含むパリティグループPG8-11とそれに逆方向に連続するパリティグループPG4-7となる。

【0042】冗長データステージング実行フラグ310がON、先読みステージング実行フラグ320が両方向と設定されている場合、ステージング範囲600は当該データD10を含むパリティグループPG8-11とそれに順方向／逆方向に連続するパリティグループPG12-15／PG4-7となる。

【0043】冗長データステージング実行フラグ310がON、先読みステージング実行フラグ320が先読みなしと設定されている場合、ステージング範囲600は当該データD10を含むパリティグループPG8-11のみとなる。

【0044】冗長データステージング実行フラグ310がOFF、先読みステージング実行フラグ320が順方向と設定されている場合、ステージング範囲600は当該データD10とそれに順方向に連続する順方向先読み量330分のデータストライプD11～D14となる。

【0045】冗長データステージング実行フラグ310がOFF、先読みステージング実行フラグ320が逆方向と設定されている場合、ステージング範囲600は当該データD10とそれに逆方向に連続する逆方向先読み量340分のデータストライプD9～D6となる。

【0046】冗長データステージング実行フラグ310がOFF、先読みステージング実行フラグ320が両方向と設定されている場合、ステージング範囲600は当該データD10とそれに順方向に連続する順方向先読み量330分のデータストライプD11～D14、およ

び、逆方向に連続する逆方向先読み量340分のデータストライプD9～D6となる。

【0047】冗長データステージング実行フラグ310がOFF、先読みステージング実行フラグ320が先読みなしと設定されている場合、ステージング範囲600は当該データD10のみとなる。

【0048】図6の中では、各組み合わせにおいて斜線で示したデータストライプがステージング範囲600である。

10 【0049】図7は、データ復元処理（ステップ408）を説明したフローチャートである。

【0050】ステージング処理の終了判定（ステップ407）において、当該リードデータストライプD10の読み出しに失敗した場合、冗長データから当該データD10を復元するデータ復元処理（ステップ408）が必要となる。

20 【0051】データ復元処理（ステップ408）では、まず、冗長データステージング処理の実行状態を確認する（ステップ701）。ここで、冗長データステージング処理を未実行の場合には冗長データステージング処理を実行し、当該データD10を復元するのに必要な同一パリティグループPG8-11の他のデータストライプD8／D9／D11と冗長データストライプP8-11がキャッシュメモリ120にステージングする（ステップ702）。また、冗長データステージング処理が実行中でまだ終了していない場合には冗長データステージング処理の終了を待つ（ステップ703）。冗長データのステージングが完了した後、データ復元手段150を用いて当該データD10を復元してキャッシュメモリ120に格納する（ステップ704）。

30 【0052】本実施例においてユーザが冗長データステージング実行を指示した場合の応答時間の短縮効果について説明する。

【0053】従来方式では当該データD10のステージング実行時に冗長データステージングを実行していない。そのため、当該データD10のステージング処理失敗が判明した後、改めてデータ復元に必要な同一パリティグループPG8-11の他のデータD8／D9／D11と冗長データP8-11のステージングを実行してから、データ復元をおこなっていた。すなわち、従来方式における当該データD10のステージングに失敗した場合の応答時間は、
従来方式での応答時間＝当該データステージング時間＋冗長データステージング時間＋データ復元時間＋その他時間となっていた。

50 【0054】これに対し、本発明の方式で、ユーザが上位装置10やユーザインタフェース手段180を用いて冗長データステージング実行フラグ310にONを設定した場合には、当該データD10ステージング時に同時

に冗長データステージングも行う（ステップ503、または、ステップ504）ため、データ復元処理（ステップ408）時には冗長データステージングが不要となり、キャッシュメモリ120上の冗長データを用いて速やかにデータ復元が行える。そのため、応答時間は、本発明の方式での応答時間＝当該データステージング時間＋データ復元時間＋その他時間

となり、従来方式に比べて、冗長データステージング時間の分だけ応答時間を短縮する事ができる。一般に、記憶媒体200からのステージングは、キャッシュメモリ120からの読み出しに比べて非常に遅く、本発明により応答時間を大きく短縮する事ができる。

【0055】次に、本実施例においてユーザが先読みステージング実行を指示した場合のシーケンシャルリード処理の高速化に対する効果について説明する。

【0056】ユーザが上位装置10やユーザインタフェース手段180を用いて先読みデータステージング実行フラグ320に順方向を設定した場合には、当該データD10をステージングする時にD10に順方向に連続する領域（D11以降）のデータも同時にステージングする。また、先読みデータステージング実行フラグ320に逆方向を設定した場合には、当該データD10をステージングする時にD10に逆方向に連続する領域（D9以前）のデータも同時にステージングし、先読みデータステージング実行フラグ320に両方向を設定した場合には、当該データD10をステージングする時にD10に順方向／逆方向に連続する領域（D11以降、および、D9以前）のデータも同時にステージングする。

【0057】このとき、従来行っていたシーケンシャル判定は不要となるため、従来方式のシーケンシャル判定に要していた時間が削減できる。

【0058】また、上位装置10から順方向／逆方向のシーケンシャルなリード要求が発行される場合、1回目のリード要求時に2回目以降のリード要求の領域も先にステージングされているため、2回目以降のリード処理ではキャッシュメモリからホストへデータ転送するだけでよいので処理時間を短縮することができる。一般に、記憶媒体200からのステージングにくらべてキャッシュメモリ120からの読み出しは非常に早いので、本発明によりシーケンシャルリード処理を大きく高速化できる。

【0059】最後に、本実施例においてユーザが先読みステージング実行を指示した場合の複数多重シーケンシャルリード処理方式、および、効果について説明する。

【0060】本発明では、シーケンシャルリード要求の多重数により制限される情報を用いないため、シーケンシャルリード要求の多重度を意識することなく処理が可能である。複数多重シーケンシャルリード要求の場合でも、上位装置10からのリード要求時にステージング制御情報テーブル300の設定内容にしたがって記憶媒体

200へステージング要求を発行する。ここで、下位へのI/O要求を管理するために一般的に使用されている、キューイング機能を持つ記憶媒体200やキューイング機能を持つ下位接続手段170を使用することにより、多重に発行したステージング要求を管理／実行することが可能となる。

【0061】したがって、従来は一般にシーケンシャル判定により先読みステージングを行っているためシーケンシャル判定を行うためのテーブルのセット数により先読みステージング処理の多重度が制限されていたが、本発明ではシーケンシャル判定が不要なためシーケンシャルリード処理の多重度に関する制限がなくなり、何多重でも先読みステージング処理を実行できるため、多数多重したシーケンシャル要求に対しても高速に処理することができる。

【0062】本発明によれば、上位装置10やユーザインタフェース手段180を通して設定することによって、従来方式のステージング、冗長データも含めたステージング、順方向／逆方向の先読みステージングの中から、使用するシステムにあわせたステージング方式を選択しチューニングすることができる。また、RAIDグループやLUなどで外部記憶サブシステム100が領域分割されている場合や、複数ホストが接続されている場合には、その領域分割単位やホスト単位にステージング制御情報テーブル300を持つことによって、領域分割毎、ホスト毎のステージング方法のチューニングも可能となる。

【0063】

【発明の効果】本発明によれば、冗長データによるデータ復元手段を持つ外部記憶サブシステムにおいて記憶媒体からのデータ読出しに失敗した時に速やかにデータ復元をおこなえるため、応答時間を短くでき最大応答時間保証に効果がある。

【0064】また、先読みステージング処理実行／先読み方向などの先読みに関する情報を保持することにより、上位装置からのシーケンシャルなデータ読み出しに対してシーケンシャル判定が不要になるため、処理時間を短縮でき、単一および複数多重のシーケンシャルな読み出しの処理高速化に効果がある。

【0065】さらに、シーケンシャル判定情報を保存しておく必要がなくなるため、シーケンシャル判定情報の数による先読み多重数の制限がなくなる。したがって、何多重にも先読み処理を動作させることができるため、多重シーケンシャル読み出しの高速化に効果がある。また、逆方向への先読みを実施することにより、データの逆再生時にも高速な処理が可能となる。

【図面の簡単な説明】

【図1】本発明による外部記憶サブシステムの構成を示す図である。

【図2】外部記憶サブシステムの記憶媒体内部のデータ

／冗長データの配置の 1 例を示す概念図である。

【図 3】本発明による外部記憶サブシステムで用いられるステージング制御情報テーブルの例を示す図である。

【図 4】本発明による外部記憶サブシステムの I/O 処理のフローチャートである。

【図 5】ステージング処理を詳細に示すフローチャートである。

【図 6】ステージング制御情報テーブルの設定値とステージング実行範囲との関係の例を示す概念図である。

【図 7】データ復元処理を示すフローチャートである。 10

【符号の説明】

10：上位装置

* 100：外部記憶サブシステム

110：制御手段

130：データ転送手段

段

150：データ復元手段

170：下位接続手段

ース手段

200：記憶媒体

300：ステージング制御情報テーブル

310：冗長データステージング実行フラグ

320：先読みステージング実行フラグ

* 330：順方向先読み量 340：逆方向先読み量

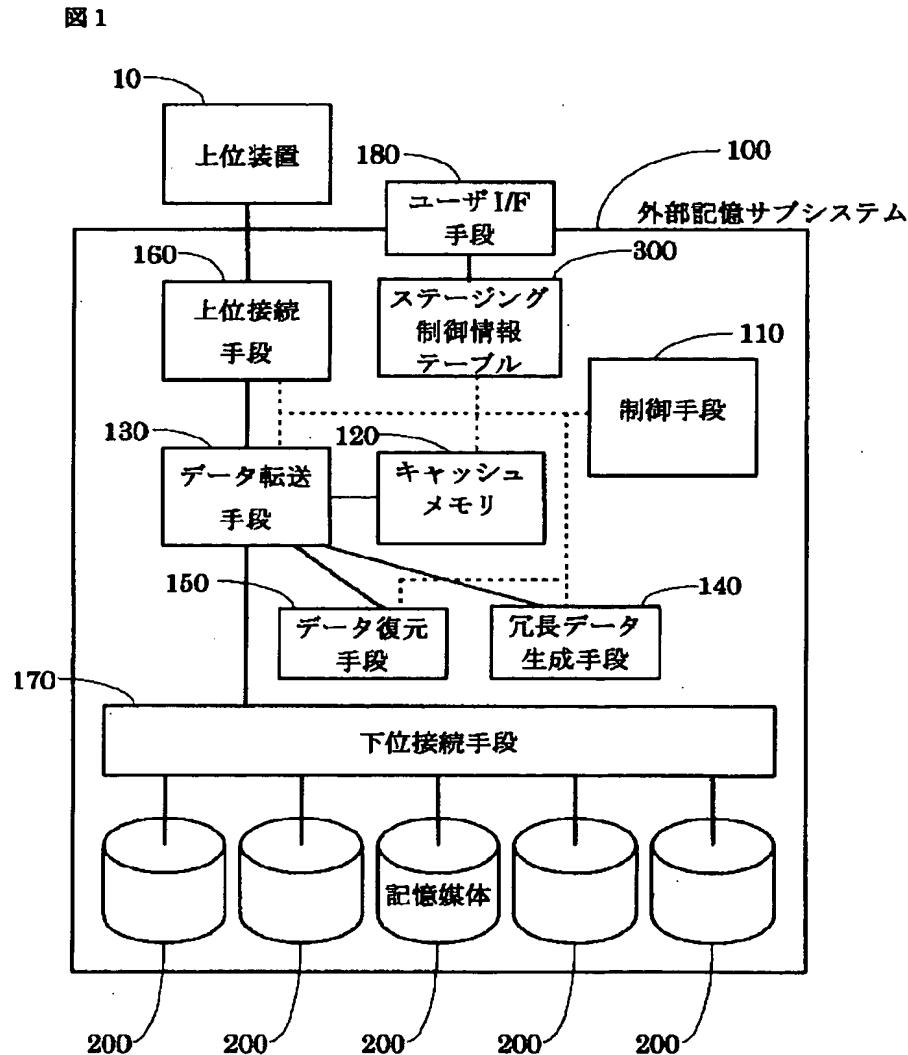
120：キャッシュメモリ

140：冗長データ生成手段

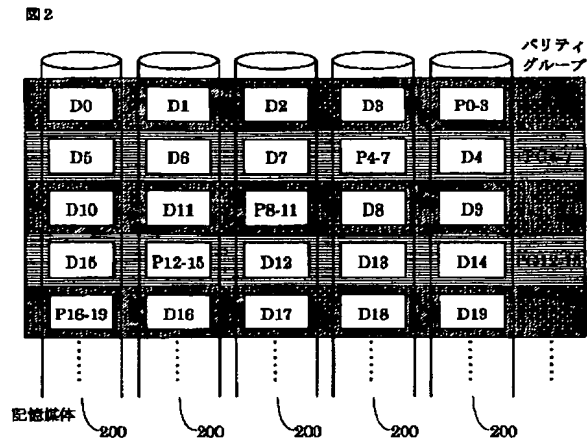
160：上位接続手段

180：ユーザインタフェース手段

【図 1】

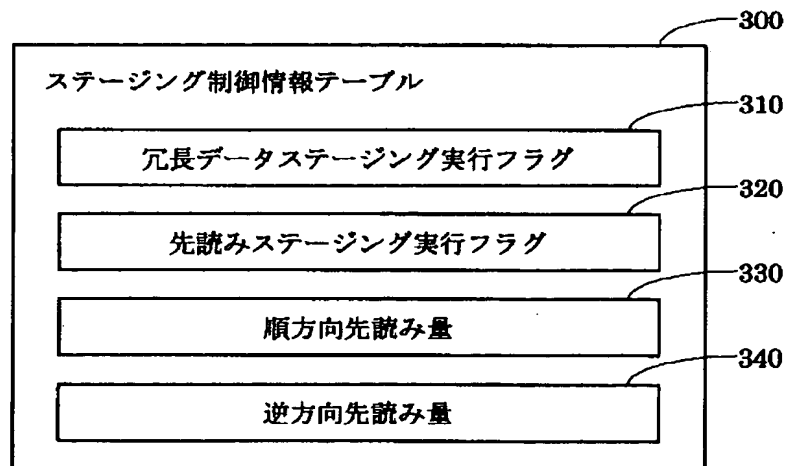


【図 2】



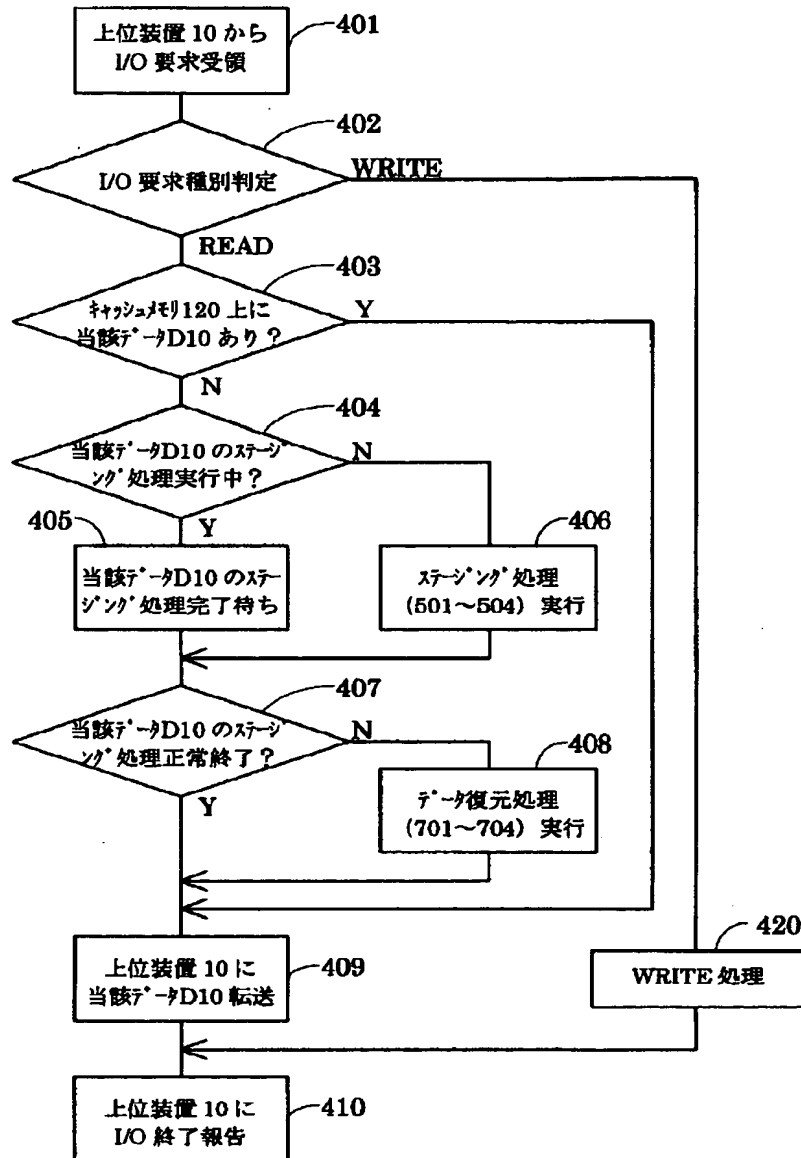
【図 3】

図 3



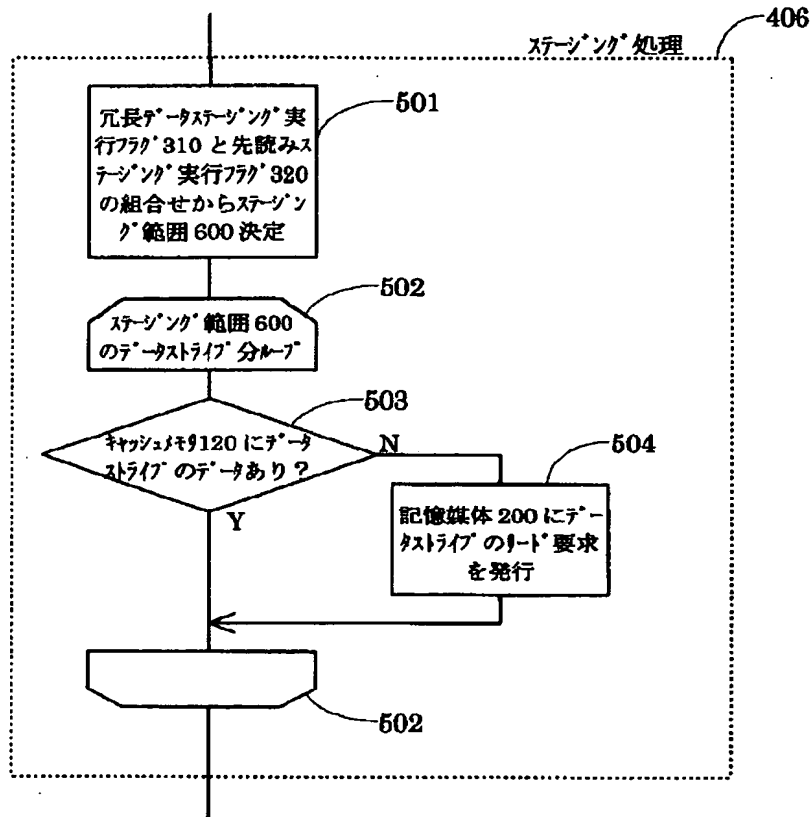
【図 4】

図 4



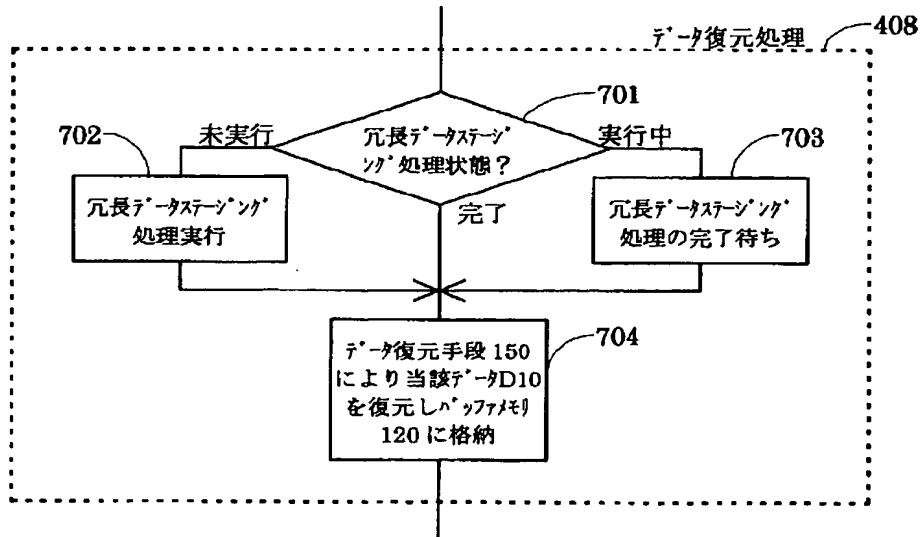
【図 5】

図 5



【図 7】


図 7



【図 6】

図 6

| | | 冗長データステージング実行フラグ 310 | | | | | | | | | |
|-----------------------|-------|---------------------------|------------------------|------------------------|--------------------------|--------------------------|---------------------------|---|--|---------------------------------|---------------------------------|
| | | ON | | | | | OFF | | | | |
| 先読みステージング実行フラグ 320 | 順方向 | D0 D5 P16-19 ... | D1 D6 D16 ... | D2 D7 D17 ... | D3 P4-7 D18 ... | P0-3 D4 D19 ... | D0 D5 P16-19 ... | D1 D6 D11 P12-14 D18 ... | D2 D7 P8-11 D12 D17 ... | D3 P4-7 D13 D18 ... | P0-3 D4 D14 D19 ... |
| | 逆方向 | D0 D5 P16-19 ... | D1 D6 D16 ... | D2 D7 D17 ... | D3 P4-7 D18 ... | P0-3 D4 D19 ... | D0 D5 P16-19 ... | D1 D6 D11 P12-14 D18 ... | D2 D7 P8-11 D12 D17 ... | D3 P4-7 D13 D18 ... | P0-3 D4 D14 D19 ... |
| | 両方向 | D0 D5 P16-19 ... | D1 D6 D16 ... | D2 D7 D17 ... | D3 P4-7 D18 ... | P0-3 D4 D19 ... | D0 D5 P16-19 ... | D1 D6 D11 P12-14 D18 ... | D2 D7 P8-11 D12 D17 ... | D3 P4-7 D13 D18 ... | P0-3 D4 D14 D19 ... |
| | 先読みなし | D0 D5 P16-19 ... | D1 D6 D16 ... | D2 D7 D17 ... | D3 P4-7 D18 ... | P0-3 D4 D19 ... | D0 D5 P16-19 ... | D1 D6 D11 P12-14 D18 ... | D2 D7 P8-11 D12 D17 ... | D3 P4-7 D13 D18 ... | P0-3 D4 D14 D19 ... |

※斜線部  がキャッシュメモリ 120 へのステージング範囲 600。